Predicting Progression to Dementia in Elderly Subjects with Mild Cognitive Impairment Using Both Cognitive and Neuroimaging Predictors

Frédéric Peters*, Sylvia Villeneuve* and Sylvie Belleville*,b,*

*a Institut Universitaire de Gériatrie de Montréal, Montréal, Canada
*b Department of Psychology, Université de Montréal, Montréal, Canada

Accepted 8 July 2013

Abstract. The objective of this work was to assess the predictive accuracy of targeted neuroimaging and neuropsychological measures for the detection of incipient dementia in individuals with mild cognitive impairment (MCI), and to examine the potential benefit of combining both classes of measures. Baseline MRI measures included hippocampal volume, cortical thickness, and white matter hyperintensities. Neuropsychological assessment focused on different aspects of episodic memory (i.e., familiarity, free recall, and associative memory) and executive control functions (i.e., working memory, switching, and planning). Global and regional cortical thinning was observed in MCI patients who progressed to dementia compared to those who remained stable, whereas no differences were found between groups for baseline hippocampal volume and white matter hyperintensities. The strongest neuroimaging predictors were baseline cortical thickness in the right anterior cingulate and middle frontal gyri. For cognitive predictors, we found that deficits in both free recall and recognition episodic memory tasks were highly suggestive of progression to dementia. Cortical thinning in the right anterior cingulate gyrus, combined to controlled and familiarity-based retrieval deficits, achieved a classification accuracy of 87.5%, a specificity of 90.9%, and a sensitivity of 83.3%. This predictive model including both classes of measures provided more accurate predictions than those based on neuroimaging or cognitive measures alone. Overall, our findings suggest that detecting preclinical Alzheimer’s disease is probably best accomplished by combining complementary information from targeted neuroimaging and cognitive classifiers, and highlight the importance of taking into account both structural and functional changes associated with the disease.

Keywords: Alzheimer’s disease, cortical thickness, episodic memory, mild cognitive impairment

INTRODUCTION

Early and accurate diagnosis of Alzheimer’s disease (AD) has become a major clinical and public health concern with the aging of the population, but also in light of the promising development of effective treatment. In recent years, substantial progress has been made in the characterization and the detection of preclinical AD, especially with the development of neuroimaging techniques, such as molecular [e.g., Pittsburgh compound B (PIB)], structural [e.g., magnetic resonance imaging (MRI)], and functional [e.g., fluorodeoxyglucose-positron emission tomography (FDG-PET)] brain imaging. It is now known that the neurodegenerative process associated with AD begins several years before the clinical manifestations of the disease, and that several neuroimaging measures derived from these techniques may be helpful in predicting individuals who later progress to AD and those who remain non-demented (see [1] for a review).
Neuromaging studies using MRI techniques have provided important insights into the pattern of structural brain damage (e.g., cortical atrophy or white matter changes) associated with progression to AD. The neuronal loss seen in prodromal AD patients is usually most prominent in the entorhinal cortex, the hippocampus, the anterior and posterior cingulate, and the frontal and temporal lobes [2–5]. Accordingly, follow-up studies that looked at grey matter atrophy or cortical thinning in these regions have found an association with increased risk of developing AD [4, 6]. Prediction methods based on neuronal loss estimates may reach a high specificity, in that they minimize the number of false positives. However, they are often limited by a lack of sensitivity, resulting in a high proportion of false negative results [7].

Neuromaging markers have proved to be particularly useful in the detection of prodromal AD patients among individuals who are characterized by mild cognitive impairment (MCI). However, studies investigating the predictive validity of neuromaging measurements have rarely included measures of cognitive functioning in predictive models and those that did included very few measures. Yet, recent findings indicate that the pattern of cognitive deficits in MCI subjects who progress to dementia is distinct from the one observed in MCI subjects who remain stable, and that specific measures of cognitive functioning may serve as reliable predictors of subsequent AD [8–10].

Memory deficits in prodromal AD appear to be particularly severe and to encompass both the familiarity and recollective aspects of recognition memory, whereas familiarity-based recognition is relatively spared in stable MCI subjects [11, 12]. In addition, there is evidence to suggest that extra-memory cognitive impairments, particularly executive dysfunction, are more common in preclinical stages of AD [10, 13–15]. Executive dysfunction may be apparent on tasks that require concurrent mental manipulation of information (e.g., working memory), mental shifting between different task-sets, planning, or decision-making [16–18].

The general aim of this study was to determine whether the combined use of structural brain measures and targeted neuropsychological assessment may improve the sensitivity and the accuracy in identifying prodromal AD. Baseline MRI and cognitive measures were collected in a cohort of individuals with MCI. A 2-years clinical follow-up allows us to determine whether these patients subsequently convert to dementia within this period. Grey matter measures included hippocampus volumetry and cortical thickness in a set of a priori region of interest. From previous neuromaging studies examining cortical thickness of prodromal AD, we know that between-group comparison yield a consistent pattern of cortical thinning in progressive-MCI encompassing the parahippocampal gyrus, anterior and posterior cingulate, middle and superior frontal areas as well as lateral temporal gyri [3, 4, 6, 19]. Baseline neuropsychological evaluation examined various aspects of episodic memory (e.g., immediate and delayed free recall, familiarity-based recognition memory, and associative memory) and targeted executive functions (e.g., resistance to interference in working memory, planning of mental operations, and task-set shifting). Furthermore, because cerebral small-vessel disease has been linked with impaired executive functions, we also included a measure of white matter hyperintensities (WMH), which is believed to be a reliable proxy of white matter ischemia [20–22]. Logistic regression analyses were used to test the ability of baseline neuropsychological and MRI measures to predict conversion from MCI to dementia and to investigate the potential improvement of combining the use of both types of measures in predictive models. Indeed, MRI markers of AD have been mostly examined in isolation, and it is important from a clinical point of view to determine whether the inclusion of neuropsychological assessment targeting cognitive dysfunction at prodromal stages of AD significantly improves the detection of progressive MCI subjects.

METHODS

Participants

A group of 45 individuals fulfilling clinical criteria for MCI and a group of 20 healthy elderly controls (EC) participate to this study. Criteria for MCI [23] included (a) cognitive complaint, preferably corroborated by an informant; (b) performance at least 1.5 standard deviation below age and education normative values on clinical tests assessing memory, language, or attentional domains; (c) essentially preserved activities of daily living; and (d) no dementia. The diagnosis of MCI was established in memory disorders clinics belonging to tertiary referral centers in Montreal, based on a semi-structured medical, neurological, neuropsychological, and clinical assessment. Experimental testing and image acquisition was performed shortly after it was established that participants met criteria for MCI. Control participants were recruited from a pool of volunteers living in the community in Montreal. They underwent the same clinical and neuropsychological
with an adapted-Peterson procedure, switching capac-
sored resistance to interference in working memory
used to assess specific executive functions. We mea-
with a word-pair learning task. Another set of tasks was
recognition task, and associative memory was tested
explicit episodic memory, familiarity-based recog-
tasks were used for measuring retrieval aspects of
control processes. Immediate and delayed free recall
iment, 1994) following a clinical assessment by an
criteria for dementia (American Psychiatric Associ-
dementia was determined by meeting the Diagnostic
annual follow-up period. Progression to
eal inclusion criteria included alcoholism, general anes-
procedure constructed and validated by Van der Linden
portion AB of the AB/AC procedure) developed by
Villemenue and Belleville [28]. Participants were pre-
and concrete words taken from the computerized
immediate tasks assessing verbal memory and executive
evaluation and were only included in the study if
they performed within normal range on clinical cog-
te tests (score that did not reach the cut-off of
Mini-Mental State Evaluation (MMSE; [24]) and the
 Mattis Dementia Rating Scale (MDRS; [25]) were
used to assess global cognitive functioning. Each con-
trol participant performed above the cut-off score
of 26 at the MMSE and 130 at the Mattis scale. The
group of MCI participants had a significantly
lower score than the control group on MMSE (MCI
group mean: 27.6 (22–29), NC group mean: 29.6
(29–30), t(1, 58) = 5.13, p < 0.05) and MDRS scale
(MCI group mean: 134.4 (114–144), NC group mean:
140.3 (133–144), t(1, 58) = 3.75, p < 0.05), whereas
no significant differences were found between the
groups when we performed t-tests comparing age
(t(1, 58) = 0.22, p > 0.824) and educational level (t(1,
58) = 0.25, p > 0.803). All participants were monitored
yearly over a 2-year follow-up period. Progression to
dementia was determined by meeting the Diagnostic
and Statistical Manual of Mental Disorders (DSM)-IV
criteria for dementia (American Psychiatric Associ-
ation, 1994) following a clinical assessment by an
experienced neurologist or geriatrician blind to the
experimental cognitive measures used in this study.

For both groups, exclusion criteria included alco-
holism, general anesthesia in the last 6 months, and
presence or history of severe psychiatric disorder, intel-
lectual deficiency, neurological diseases or events (e.g.,
stroke, Parkinson’s disease, epilepsy, brain anoxia),
psychiatric disorder (schizophrenia, major depression,
alcoholism), traumatic brain injury, or a systemic dis-
ease known to impair cognition. All participants were
native French-speakers and had normal or corrected
vi-sion. In the learning phase, participants were asked
to read the words and tried to memorize them. In the
retrieval phase, the first word of the pair was presented
as a cue and the participants had to recall the second
word of the pair. We measured the proportion of pairs
correctly recalled.

Working memory
Working memory was measured with an adapted
Brown-Peterson procedure, taken from the computer-
ized Memoria Battery [26]. Consonant trigrams were
presented auditorily to participants and participants
were asked to recall series these trigrams after differ-
ent time delays (10, 20, or 30 seconds). During the
delay, participants were asked to add one to a series of
randomly presented numbers presented orally by the
participants.

Neuropsychological assessment

The neuropsychological testing consisted of exper-
imental tasks assessing verbal memory and executive
control processes. Immediate and delayed free recall
tasks were used for measuring retrieval aspects of
explicit episodic memory, familiarity-based recog-
nition memory were evaluated via a forced-choice
recognition task, and associative memory was tested
with a word-pair learning task. Another set of tasks was
used to assess specific executive functions. We mea-
sured resistance to interference in working memory
with an adapted-Peterson procedure, switching capac-
ties with the Trail Making Test, and planning abilities
using the Tower of London.

Immediate free recall and recognition task

Participants were asked to study a list of 16 fre-
quent and concrete words taken from the computerized
Memoria Battery [26]. Words were presented visu-
ally. Free recall and recognition were measured shortly
after the learning phase. In the recognition condition,
each target word was presented with a distractor that
belongs to the same semantic category and was com-
parable in terms of lexical frequency and word length.
We measured the number of words correctly recalled
for the free recall condition, and the number of correct
responses for the recognition condition.

Delayed verbal free recall memory

We used an adaptation of Grober and Buschke’s pro-
cedure constructed and validated by Van der Linden
et al. [27]. It consists of 16 words belonging to differ-
ent semantic categories printed on four cards with four
items per card. One four-item card was presented at a
time. In the learning phase, participants were asked to
point to and say the relevant word (e.g., celery) when
the experimenter named its semantic category (e.g.,
vegetable). Free recall procedure was carried out after
delay of 20 minutes. We measured the difference
between the number of words correctly recalled at the
end of the learning phase and after the delay.

Word-pair learning

This task is an adaptation of the verbal learning task
(portion AB of the AB/AC procedure) developed by
Villemenue and Belleville [28]. Participants were pre-
sented with a list of 12 semantically related words
pairs. The words were mono or bisyllabic and were
of high lexical frequency [29]. Participants were asked
to read the words and tried to memorize them. In the
retrieval phase, the first word of the pair was presented
as a cue and the participants had to recall the second
word of the pair. We measured the proportion of pairs
correctly recalled.
Task switching

We measured task switching abilities using the Trail Making Test [30]. There are two parts to the test in which subjects are required to connect in sequential order the dots of 25 consecutive targets on a sheet of paper. In the part «A», the targets are all numbers (1, 2, 3, etc.), whereas in the part «B», the subject have to alternate between letters and numbers in ascending order (1, A, 2, B, etc.). We calculated the proportional cost in completion time for the part «B» (i.e., with switching) as compared to the part «A» (i.e., without switching).

Executive planning

Planning abilities were tested with Tower of London task [31]. This task consists of two boards with three pegs. On one, the examiner places three colored wooden balls (blue, green, and red) in a goal position by observing two rules: only one ball may be moved at a time, and only a specified number of balls may be left on each peg at a time. In presence of rule breaking, the examiner stopped the test to remind the subject that a rule has been broken. We measured the breaking, the examiner stopped the test to remind the subject that a rule has been broken. We measured the

MRI acquisition and image processing

Participants were scanned using a 3 T Siemens TIM MRI scanner at the Functional Neuroimaging Unit of the Centre de Recherche de l’Institut Universitaire de Gériatrie de Montréal using a three-dimensional T1-weighted gradient-echo sequence (TR/TE/TI: 2300/2.91/900 ms, flip angle: 9°, 160 slices, field of view: 256 × 240 mm, matrix: 256 × 240, voxel size: 1 × 1 × 1 mm, 12-channels coil).

Manual measure of hippocampal volume

The hippocampal volume was obtained following a manual tracing method, using the Anatomist/BrainVISA 3.1 package (http://brainvisa.info/). Four reference points were first positioned (anterior commissure, posterior commissure, interhemispheric plan, and left hemisphere) to allow volume re-orientation and to generate a transformation referenced to Talairach atlas. This transformation was used to reduce the volumes of different subjects to the same reference volume without altering the data. Included in our hippocampus measurements were the cornu ammonis (CA1–CA4), the dentate gyrus, the subiculum, and the alveus, based on the protocol described in Wu et al. [32]. Rostrally, the portion of the uncus connecting to the amygdala was included. Caudally, the fimbria (white matter fibers connecting the hippocampus to the fornix) was excluded from the measurements. The sagittal orientation was initially used, and subsequent corrections were made to the coronal orientation as necessary. Left and right volumes were calculated separately. Intracranial volume (ICV) was measured following the procedure of Eritaia et al. [33]. Hippocampal volumes were normalized to head size using the formula: (Hippocampal volume/ICV). All volumes were measured by a single experienced rater, blind to participant diagnosis.

Cortical thickness

Every T1 weighted volume was processed through the CIVET pipeline [34] using the following sequence of automated treatments: 1) Correction for non-uniformity intensity artifacts; 2) Linear registration of the native MRI into a Talairach like stereotaxic space based on the ICBM152 template [35–37]; 3) Segmentation of the resulting volumes into white matter (WM), grey matter (GM), cerebrospinal fluid (CSF), and background using an advanced neural net classifier [38], and partial volume estimation [39] to attribute a continuous classification value to each voxel of the volume; 4) Fitting of the GM and WM surfaces using deformable models [40, 41], resulting in two 40,962-point surfaces; 5) Computation of the cortical thickness at each vertex as the Euclidian distance between corresponding vertices of inner and outer surfaces [42]; and 6) Application of a 20 mm FWHM surface based smoothing kernel. A standardized cortical surface object was computed from averaging the 60 normalized surfaces of the participants, to adequately display statistical results. For predictive analyses, mean cortical thickness was computed on a 10 mm radius sphere, around the averaged coordinates published for the corresponding location of interest. The following a priori locations of interest were used, based on published coordinates in the previous studies directly comparing stable MCI patients and progressive MCI patients.
and using similar image processing methods [3, 4]: bilateral parahippocampal gyrus (MNI coordinates: 33, −20, −25, −29, 8, −45), posterior cingulate (6, −39, 43; −3, −28, 32), precuneus (6, −43, 45; −5, −76, 46), middle frontal gyrus (40, 25, 36; −41, 35, 17), superior frontal gyrus (5, 65, 15; −5, 29, 39), and inferior temporal gyrus (46, −25, −24; −41, −19, −28), as well as the right anterior cingulate (12,41,18), the left middle temporal gyrus (−64, −51, −2), and the right superior temporal gyrus (64, −16, −12).

White matter hyperintensity rating scale

The Age-Related White Matter Changes rating scale of the European Task Force was used as a general measure of WMH severity [43]. This scale uses a four-point severity rating (0-normal; 1-punctate; 2- beginning confluence; 3-diffuse involvement) for each of five brain regions (i.e., frontal area, parieto-occipital area, temporal area, infratentorial area, and basal ganglia) in the right and left hemispheres separately. Ratings were done on MRI images on computer screen with T2 and FLAIR images. Total score ranges from 0 to 30. White matter changes on MRI were defined with T2 and FLAIR images. Total score ranges from 0 to 30. White matter changes on MRI were defined with T2 and FLAIR images. Total score ranges from 0 to 30.

Statistical analysis

First, between-group comparisons were performed on baseline measures. One-way analyses of variance (ANOVA) followed by Tukey’s post hoc tests were performed on demographic characteristics, neuropsychological measures, hippocampal volume, WMH volume, and average thickness across the entire cortex. Between-group comparisons for gender distribution were analyzed by means of χ² tests. For local differences in cortical thickness, all statistical analyses were conducted using the SurfStat toolbox (http://www.math.mcgill.ca/keith/surfstat/) based on Random Field Theory (RFT) [44], for Matlab (R2008b, The Mathworks, Natick, MA, USA). Age, gender, and educational level were used as confounding covariates. Clusters are reported reaching a significance level of p < 0.05 RFT corrected.

Second, predictive analyses were performed with neuropsychological and neuroimaging measures as predictors of conversion to dementia in all participants with MCI. It should be stated that for these predictive analyses, we used mean local cortical thickness measures based on coordinates published in previous studies comparing stable and progressive MCI subjects (see the previous section). Two separate forward stepwise regression analyses were first conducted for clustered neuroimaging and cognitive features (P-to-enter = 0.05). Then, both neuroimaging and cognitive measures were entered in an independent forward stepwise regression analysis. We estimated the accuracy, the sensitivity, and the specificity. Each predictive model was subjected to a leave-one-out cross-validation. This method has been shown to optimize the use of a limited dataset while preserving from over-determination, giving an almost unbiased estimate of the generalization properties of statistical models [45]. In order to further assess the significance of the observed prediction accuracy, permutation testing was used to derive a p-value. We randomly permuted the class labels 1000 times and statistical analyses were repeated using leave-one-out cross-validation in every iteration of the permutation procedure. The predictive accuracy was kept for each one of the 1000 permutation tests, then forming the distribution of the accuracy test-statistic. The P-level was computed as the fraction of permutations that gave rise to a greater value of the test statistic than the point estimate.

RESULTS

Five out of the 45 MCI participants of this study were lost to follow-up. The remaining MCI were monitored yearly to determine whether they later progressed to dementia (pMCI) or remained stable (sMCI) over a 2-years follow-up. Eighteen MCI subjects (45%) progressed to dementia, whereas 22 (55%) remained cognitively stable over successive clinical evaluations. Three out of the 18 pMCI were diagnosed as probable mixed dementia as symptoms of both AD and vascular dementia coexisted. The other 15 pMCI participants were diagnosed as probable AD patients.

Between-group comparisons of baseline measures

There were no statistical differences between elderly control (EC) and MCI groups (sMCI and pMCI) for age, educational level, and gender distribution (all p > 0.05; see Table 1). For all baseline episodic memory measures, significant differences were found between controls and subjects with MCI (immediate recall: F(2,57) = 20.5, p < 0.001; immediate recognition: F(2,57) = 19.5, p < 0.001; delayed free recall: F(2,57) = 13.1, p < 0.001; word-pair learning: (F(2,57) = 15.8, p < 0.001, see Table 2). Post hoc
Table 1

<table>
<thead>
<tr>
<th>HC</th>
<th>sMCI</th>
<th>pMCI</th>
<th>Group effect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>n</strong></td>
<td>20</td>
<td>22</td>
<td>18</td>
</tr>
<tr>
<td>Gender (m, f)</td>
<td>6, 14</td>
<td>9, 13</td>
<td>8, 10</td>
</tr>
<tr>
<td>Age (years)</td>
<td>72.0 (6.9)</td>
<td>70.4 (7.1)</td>
<td>72.9 (6.3)</td>
</tr>
<tr>
<td>Education (years)</td>
<td>13.2 (3.9)</td>
<td>13.0 (4.8)</td>
<td>13.8 (5.2)</td>
</tr>
<tr>
<td>MMSE (/30)</td>
<td>29.6 (0.5)</td>
<td>28.1 (1.4)</td>
<td>27.2 (2.0)</td>
</tr>
<tr>
<td>Mattis dementia rating scale (/142)</td>
<td>140.4 (3.0)</td>
<td>137.9 (3.7)</td>
<td>130.3 (7.6)</td>
</tr>
</tbody>
</table>

HC, group of healthy elderly controls; sMCI, group of MCI patients who remained cognitively stable after a 2-year clinical follow-up; pMCI, group of MCI patients who progressed to dementia after a 2-year clinical follow-up.

Table 2

<table>
<thead>
<tr>
<th>HC</th>
<th>sMCI</th>
<th>pMCI</th>
<th>Group effect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Neuroimaging measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normalized hippocampal volume</td>
<td>0.326 (0.038)</td>
<td>0.313 (0.037)</td>
<td>0.310 (0.043)</td>
</tr>
<tr>
<td>Wallhund scores of WMH</td>
<td>48.8 (4.2)</td>
<td>66.3 (3.7)</td>
<td>63.3 (4.2)</td>
</tr>
<tr>
<td>Mean global cortical thickness</td>
<td>2.91 (0.12)</td>
<td>2.95 (0.14)</td>
<td>2.80 (0.20)</td>
</tr>
</tbody>
</table>

HC, group of healthy elderly controls; sMCI, group of MCI patients who remained cognitively stable after a 2-year clinical follow-up; pMCI, group of MCI patients who progressed to dementia after a 2-year clinical follow-up; WMH, white matter hyperintensity.

Table 2

<table>
<thead>
<tr>
<th>HC</th>
<th>sMCI</th>
<th>pMCI</th>
<th>Group effect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognitive measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Episodic memory tasks (% of correct)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate free recall</td>
<td>53.4 (16.2)</td>
<td>44.3 (11.2)</td>
<td>27.4 (9.4)</td>
</tr>
<tr>
<td>Immediate recognition</td>
<td>96.2 (5.5)</td>
<td>94.0 (4.9)</td>
<td>80.6 (13.0)</td>
</tr>
<tr>
<td>Delayed free recall</td>
<td>75.9 (13.6)</td>
<td>64.5 (14.4)</td>
<td>48.6 (23.2)</td>
</tr>
<tr>
<td>Word-pair learning</td>
<td>63.9 (21.0)</td>
<td>47.5 (24.3)</td>
<td>25.4 (16.5)</td>
</tr>
</tbody>
</table>

HC, group of healthy elderly controls; sMCI, group of MCI patients who remained cognitively stable after a 2-year clinical follow-up; pMCI, group of MCI patients who progressed to dementia after a 2-year clinical follow-up.

There were no significant differences between the study groups for hippocampal volume, F(2,57) = 0.4, P > 0.05, or white matter hyperintensities severity, F(2,57) = 2.7, P = 0.11 (see Table 2). In contrast, the average thickness across the entire cortex was significantly thinner in the pMCI group than in both the sMCI group and the control group (all P < 0.05). Statistical differences in local cortical thickness were found between the pMCI group and the sMCI group (P ≤ 0.05 RFT corrected; see Fig. 1): in the right parahippocampal gyrus (X; Y; Z in MNI coordinates of highest peak value: 24; −11; −24), anterior cingulate gyrus (11; 45; 4), middle and superior frontal gyri (−7; −43; 12), precuneus (right: 10; −50; 24; left: −9; −59; 24), middle/posterior cingulate gyrus (right: 8; −48; 29; left: −7; −43; 41) and a bilateral region of the supramarginal gyrus in the vicinity of the Rolandic operculum (right: 51; −31; 20; left: −44; −31; 20).
Predicting conversion to dementia with regression analyses

First, separate forward stepwise regression analyses were performed with clustered sets of neuroimaging and cognitive measures as predictors of conversion to dementia in subjects with MCI. For the cognitive measures, the stepwise selection procedure identified immediate free recall and immediate recognition scores as the best combination of cognitive predictors. Overall evaluation of this model using the likelihood ratio test revealed that the model was more effective than the null hypothesis ($\chi^2 = 30.2, p < 0.001$). Both individual measures were significant predictors of conversion to AD after the 2-year follow-up (Immediate Free Recall: $\chi^2 = 6.9, p < 0.001$; Immediate Recognition: $\chi^2 = 4.8, p < 0.01$). In order to reduce overfitting, the assessment of the predictive accuracy was subjected to a leave-one-out cross-validation procedure. Overall classification accuracy of this model was 82.5%, the specificity was 90.9%, and the sensitivity was 72.2%. The positive predictive value was 80.0%, and the negative predictive value was 80.0%. The area under the curve was 0.93 at $c = 0.50$. For the analysis based on neuroimaging measures, only the right anterior cingulate and middle frontal gyri were selected by the stepwise regression approach. Likelihood ratio test for this model was significant ($\chi^2 = 14.7, p < 0.001$), and Wald chi-square statistics were marginally significant for individual local thickness measures (right anterior cingulate: $\chi^2 = 3.74, p = 0.053$; right middle frontal: $\chi^2 = 3.71, p = 0.054$). The discrimination accuracy of this subset of neuroimaging predictors was 75%, the specificity was 86.4%, and sensitivity was 61.1%. The positive predictive value was 78.6%, the negative predictive value was 73.1%, and the area under the curve was 0.81 at $c = 0.50$.

Second, an independent stepwise logistic regression analysis was performed with all cognitive and neuroimaging measures. Two cognitive measures (Immediate free recall and immediate recognition scores) and one neuroimaging measure (local cortical thickness in the right anterior cingulate gyrus) entered in the model. This composite model was significantly more effective than the null model ($\chi^2 = 39.92, p < 0.001$) and goodness-of-fit test was insignificant ($\chi^2 = 16.5, p = 0.998$), suggesting that the null hypothesis of a good model fit was tenable. Individual measures were significant predictors of conversion to AD (Immediate free recall: $\chi^2 = 6.05, p > 0.05$; right anterior cingulate thickness: $\chi^2 = 3.85, p < 0.05$) except for immediate recognition score ($\chi^2 = 3.01, p = 0.08$). Using leave-one-out cross-validation, the overall accuracy of the composite model was 87.5%, the specificity of the model was 90.9%, and its sensitivity was improved (83.3%) as compared to clustered
models (61% to 72%). The positive predictive value was 88.2%, the negative predictive value was 87.0%, and the area under the curve was 0.98 at $c=0.50$. Finally, the predictive accuracy obtained with the original diagnostic labels was significantly above chance levels estimated over 1000 random label permutations ($p<0.001$), confirming that we can reject the null hypothesis that there is no information in the data about the label being predicted.

DISCUSSION

The results of this study indicate that reliable estimates of the risk of conversion from MCI to dementia can be derived from baseline quantitative structural MRI measures and targeted neuropsychological assessment. In a series of logistic regression analyses, we showed that cortical thickness measures in several a priori locations and different memory measures were significant predictors of MCI conversion to AD or mixed dementia. Most notably, the use of predictive models including both classes of measures provided more accurate predictions than those based on neuroimaging or cognitive measures alone. More specifically, the use of the composite approach increased the level of sensitivity as compared to the separate approach, that is the likelihood of identifying patients with MCI who converted to dementia during the 2-years clinical follow-up.

As expected, comparisons between the groups on cognitive measures revealed that although memory deficits were the most prominent features in both stable and progressive MCI groups, progressive MCI subjects performed significantly worse than stable MCI subjects in all episodic memory tasks. These results support the view that episodic memory deficits in prodromal AD are more pronounced than in non-progressive MCI and encompass multiple aspects of verbal memory retrieval, including familiarity-based recognition of verbal information, which is relatively spared in stable MCI subjects [11, 12, 46]. Importantly, logistic regression analyses indicated that free recall and recognition measures were the most significant cognitive predictors of dementia in MCI subjects. Impaired controlled episodic retrieval, when associated with deficits in familiarity-based recognition were highly suggestive of progression to dementia in individuals with MCI, confirming that this profile of episodic memory deficits can be used as reliable neuropsychological markers of incipient AD [9, 47]. In contrast, although impaired working memory and planning capacities were found in the progressive MCI group, executive function measures did not add sig-
The current study examined several neuroimaging measures commonly used to detect in vivo structural brain changes associated with AD and associated disorders (i.e., hippocampal volumetry, grey matter loss, and level of white matter intensity). On a group level, significant differences between stable and progressive MCI was only observed for baseline cortical thickness measures (at both the whole-cortex and regional levels). Significant cortical thinning was present in the parahippocampal gyrus, the anterior and posterior cingulate gyrus, the middle and superior frontal gyri, the bilateral middle temporal gyrus, and the temporo-parietal junction. This pattern of cortical thinning is very similar with the one reported in previous studies of cortical thinning associated with progression from MCI to AD [3, 4, 6, 19].

For neuroimaging measures, our classification results revealed that cortical thickness in several a priori locations of interest, including areas in the entorhinal, cingulate, frontal, and temporal cortices, had significant predictive value. Using a forward stepwise selection procedure, we found that the best classification accuracy was obtained when cortical thickness measures in the right anterior cingulate and middle frontal gyri were entered as predictor variables. These results support previous studies showing that grey matter loss in the anterior cingulate and medial frontal cortex regions are among the most predictive structural brain changes of progression from MCI to AD [6, 42, 50, 51]. Interestingly, there is also evidence of metabolic decreases and amyloid-β deposition in the anterior cingulate and frontal regions in prodromal stages of AD [52-55], suggesting a specific progression of the neurodegenerative process in these brain regions for individuals with progressive MCI. Although the role of the anterior cingulate gyrus remains equivocal, this region has been proposed to be part of an executive control network that ensure conflict detection and selection of coordinated goal-directed behavior [56, 57]. Anterior cingulate dysfunction in AD has also been related to a lack of introspective knowledge and metacognitive functioning in AD [56, 58].

Importantly, we showed that cortical thinning in the right anterior cingulate gyrus, when combined to controlled recollection and familiarity-based recognition deficits, provided more accurate predictions (overall accuracy of 87.5%) than those based on cognitive or neuroimaging measures alone (82.5% and 75%, respectively). This prediction improvement was due to increased ability to detect true incipient AD patients among individual with MCI, as revealed by the sensitivity advantage (83.3%) of this composite model over models based on clustered set of cognitive or neuroimaging measures (72% and 61%, respectively). Thus, our results indicate that structural changes in the anterior cingulate and episodic memory deficits provide complementary information for the diagnostic classification. More generally, these findings reinforce the view that the detection of prodromal AD may benefit from the combination of targeted neuropsychological and neuroimaging markers [9, 59]. Current findings also provide empirical support to the most recent and prevailing diagnostic guidelines for AD[60, 61] arguing that the examination of biomarkers (including neuroimaging measures), together with clinical and neuropsychological examination, is expected to improve diagnostic accuracy.

In this study, although significant cortical thinning was reported in the parahippocampal gyrus, measures of hippocampal volume did not differ between healthy elderly, stable progressive, and non-progressive MCI subjects. It is possible that the manual hippocampus volume measurement used in this study was not able to capture the neurodegenerative changes at the prodromal stage of AD. Indeed, recent findings suggest that local hippocampus measures or hippocampus shape features may be more efficient for detecting early brain structural impairments seen in the progressive MCI group [62, 63]. Future studies should determine whether inclusion of these fine-grained hippocampus features, together with episodic memory and other structural brain changes, improve the classification of stable and progressive MCI subject. Finally, although the sample size of the MCI group in our study was relatively small, it should be noted that the cognitive and brain structural impairments seen in the progressive MCI group were quite consistent with those reported in recent studies with a larger cohort of individuals with MCI. Furthermore, prediction results were obtained after applying a leave-one-out cross-validation procedure that optimizes the use of a limited dataset while minimizing the risk of over-determination.

In conclusion, the results of this study demonstrate that both quantitative structural neuroimaging and cognitive measures may be used as valuable predictors of conversion from MCI to dementia after a period of...
ACKNOWLEDGMENTS

This research was supported by funds to SB from the Canadian Institute of Health Research (CIHR) and Heart and Stroke Foundation. FP was supported by a postdoctoral fellowship and SV by a doctoral fellowship both from CIHR. We thank Samira Melah for collecting the neuroimaging data, Danielle Tisserand for hippocampal volume measurement and Pierre Bourgois for WML quantification. We also thank Silvana Marti and Marc-Antoine Labelle for their contribution to the neuropsychological assessment.


REFERENCES


